Internet Engineering Task Force (IETF)
Request for Comments: 8227
Category: Standards Track
ISSN: 2070-1721

W. Cheng
L. Wang
H. Li
China Mobile
H. van Helvoort
Hai Gaoming BV
J. Dong
Huawei Technologies

August 2017

MPLS-TP Shared-Ring Protection (MSRP) Mechanism for Ring Topology

Abstract

This document describes requirements, architecture, and solutions for MPLS-TP Shared-Ring Protection (MSRP) in a ring topology for point-to-point (P2P) services. The MSRP mechanism is described to meet the ring protection requirements as described in RFC 5654. This document defines the Ring Protection Switching (RPS) protocol that is used to coordinate the protection behavior of the nodes on an MPLS ring.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc8227.

Table of Contents

1. Introduction

As described in Section 2.5.6.1 of [RFC5654], several service providers have expressed much interest in operating an MPLS Transport Profile (MPLS-TP) in ring topologies and require a high-level survivability function in these topologies. In operational transport network deployment, MPLS-TP networks are often constructed using ring topologies. This calls for an efficient and optimized ring protection mechanism to achieve simple operation and fast, sub 50 ms, recovery performance.

This document specifies an MPLS-TP Shared-Ring Protection mechanism that meets the criteria for ring protection and the ring protection requirements described in Section 2.5.6.1 of [RFC5654].

The basic concept and architecture of the MPLS-TP Shared-Ring Protection mechanism are specified in this document. This document describes the solutions for point-to-point transport paths. While the basic concept may also apply to point-to-multipoint transport paths, the solution for point-to-multipoint transport paths is out of the scope of this document.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Terminology and Notation

Terminology:

Ring node: All nodes in the ring topology are ring nodes, and they MUST actively participate in the ring protection.

Ring tunnel: A ring tunnel provides a server layer for the Label Switched Paths (LSPs) traversing the ring. The notation used for a ring tunnel is: R<d><p><X> where <d> = c (clockwise) or a (anticlockwise), <p> = W (working) or P (protecting), and <X> = the node name.

Ring map: A ring map is present in each ring node. The ring map contains the ring topology information, i.e., the nodes in the ring, the adjacency of the ring nodes, and the status of the links between ring nodes (Intact or Severed). The ring map is used by every ring node to determine the switchover behavior of the ring tunnels.

Notation:

The following syntax will be used to describe the contents of the label stack:

1. The label stack will be enclosed in square brackets ("[]").

2. Each level in the stack will be separated by the '|' character. It should be noted that the label stack may contain additional layers. However, we only present the layers that are related to the protection mechanism.

3. If the label is assigned by Node X, the Node Name is enclosed in parentheses ("()").

3. MPLS-TP Ring Protection Criteria and Requirements

The generic requirements for MPLS-TP protection are specified in [RFC5654]. The requirements specific for ring protection are specified in Section 2.5.6.1 of [RFC5654]. This section describes how the criteria for ring protection are met:

a. The number of Operations, Administration, and Maintenance (OAM) entities needed to trigger protection

   Each ring node requires only one instance of the RPS protocol per ring. The OAM of the links connected to the adjacent ring nodes has to be forwarded to only this instance in order to trigger protection. For detailed information, see Section 5.2.

b. The number of elements of recovery in the ring

   Each ring node requires only one instance of the RPS protocol and is independent of the number of LSPs that are protected. For detailed information, see Section 5.2.

c. The required number of labels required for the protection paths

   The RPS protocol uses ring tunnels, and each tunnel has a set of labels. The number of ring tunnel labels is related to the number of ring nodes and is independent of the number of protected LSPs. For detailed information, see Section 4.1.2.

d. The amount of control and management-plane transactions

   Each ring node requires only one instance of the RPS protocol per ring. This means that only one maintenance operation is required per ring node. For detailed information, see Section 5.2.

e. Minimize the signaling and routing information exchange during protection

   Information exchange during a protection switch is using the in-band RPS and OAM messages. No control-plane interactions are required. For detailed information, see Section 5.2.

4. Shared-Ring Protection Architecture

4.1. Ring Tunnel

This document introduces a new logical layer of the ring for shared-ring protection in MPLS-TP networks. As shown in Figure 1, the new logical layer consists of ring tunnels that provide a server layer for the LSPs traversing the ring. Once a ring tunnel is established, the forwarding and protection switching of the ring are all performed at the ring tunnel level. A port can carry multiple ring tunnels, and a ring tunnel can carry multiple LSPs.

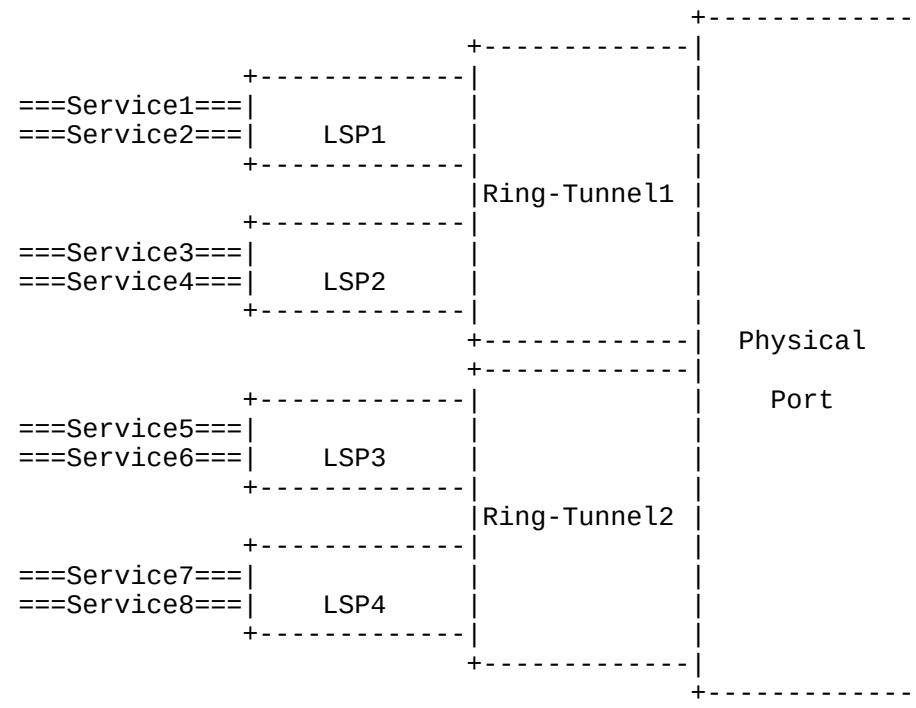

Figure 1: The Logical Layers of the Ring

The label stack used in the MPLS-TP Shared-Ring Protection mechanism is [Ring Tunnel Label|LSP Label|Service Label](Payload) as illustrated in Figure 2.

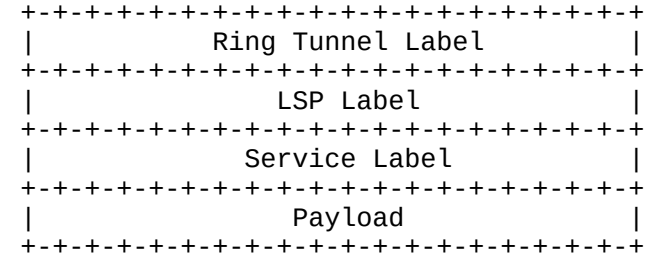

Figure 2: Label Stack Used in MPLS-TP Shared-Ring Protection

4.1.1. Establishment of the Ring Tunnel

The Ring tunnels are established based on the egress nodes. The egress node is the node where traffic leaves the ring. LSPs that have the same egress node on the ring and travel along the ring in the same direction (clockwise or anticlockwise) share the same ring tunnels. In other words, all the LSPs that traverse the ring in the same direction and exit from the same node share the same working ring tunnel and protection ring tunnel. For each egress node, four ring tunnels are established:

- one clockwise working ring tunnel, which is protected by the anticlockwise protection ring tunnel
- one anticlockwise protection ring tunnel
- one anticlockwise working ring tunnel, which is protected by the clockwise protection ring tunnel
- one clockwise protection ring tunnel

The structure of the protection tunnels is determined by the selected protection mechanism. This will be detailed in subsequent sections.

As shown in Figure 3, LSP1, LSP2, and LSP3 enter the ring from Node E, Node A, and Node B, respectively, and all leave the ring at Node D. To protect these LSPs that traverse the ring, a clockwise working ring tunnel (RcW_D) via E->F->A->B->C->D and its anticlockwise protection ring tunnel (RaP_D) via D->C->B->A->F->E->D are established. Also, an anticlockwise working ring tunnel (RaW_D) via C->B->A->F->E->D and its clockwise protection ring tunnel (RcP_D) via D->E->F->A->B->C->D are established. For simplicity, Figure 3 only shows RcW_D and RaP_D. A similar provisioning should be applied for any other node on the ring. In summary, for each node in Figure 3, when acting as an egress node, the ring tunnels are created as follows:

- To Node A: RcW_A, RaW_A, RcP_A, RaP_A
- To Node B: RcW_B, RaW_B, RcP_B, RaP_B
- To Node C: RcW_C, RaW_C, RcP_C, RaP_C
- To Node D: RcW_D, RaW_D, RcP_D, RaP_D
- To Node E: RcW_E, RaW_E, RcP_E, RaP_E
- To Node F: RcW_F, RaW_F, RcP_F, RaP_F

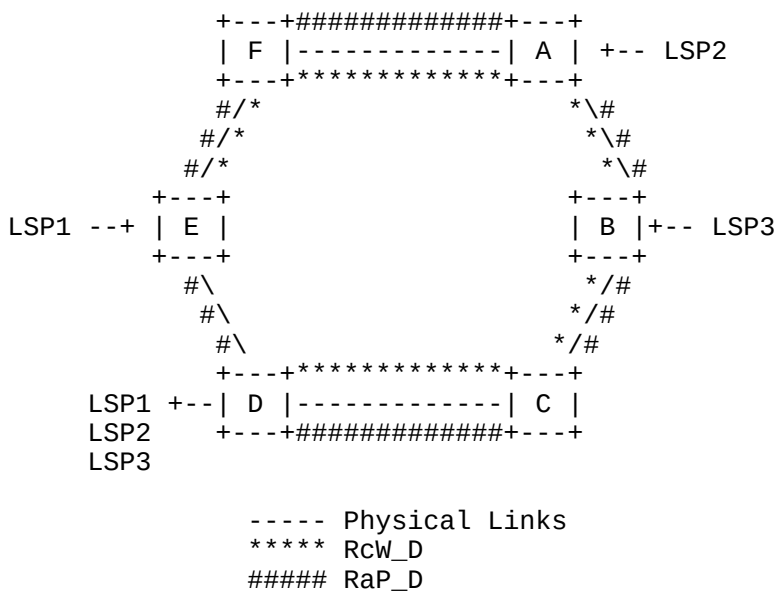

Figure 3: Ring Tunnels in MSRP

Through these working and protection ring tunnels, LSPs that enter the ring from any node can reach any egress nodes on the ring and are protected from failures on the ring.

4.1.2. Label Assignment and Distribution

The ring tunnel labels are downstream-assigned labels as defined in [RFC3031]. The ring tunnel labels on each hop of the ring tunnel can be either configured statically, provisioned by a controller, or distributed dynamically via a control protocol. For an LSP that traverses the ring tunnel, the ingress ring node and the egress ring node are considered adjacent at the LSP layer, and LSP label needs to be allocated at these two ring nodes. The control plane for label distribution is outside the scope of this document.

4.1.3. Forwarding Operation

When an MPLS-TP transport path, i.e., an LSP, enters the ring, the ingress node on the ring pushes the working ring tunnel label that is used to reach the specific egress node and sends the traffic to the next hop. The transit nodes on the working ring tunnel swap the ring tunnel labels and forward the packets to the next hop. When the packet arrives at the egress node, the egress node pops the ring tunnel label and forwards the packets based on the inner LSP label

and service label. Figure 4 shows the label operation in the MPLS-TP Shared-Ring Protection mechanism. Assume that LSP1 enters the ring at Node A and exits from Node D, and the following label operations are executed.

1. Ingress node: Packets of LSP1 arrive at Node A with a label stack [LSP1] and are supposed to be forwarded in the clockwise direction of the ring. The label of the clockwise working ring tunnel RcW_D will be pushed at Node A, the label stack for the forwarded packet at Node A is changed to [RcW_D(B)|LSP1].

2. Transit nodes: In this case, Nodes B and C forward the packets by swapping the working ring tunnel labels. For example, the label [RcW_D(B)|LSP1] is swapped to [RcW_D(C)|LSP1] at Node B.

3. Egress node: When the packet arrives at Node D (i.e., the egress node) with label stack [RcW_D(D)|LSP1], Node D pops RcW_D(D) and subsequently deals with the inner labels of LSP1.

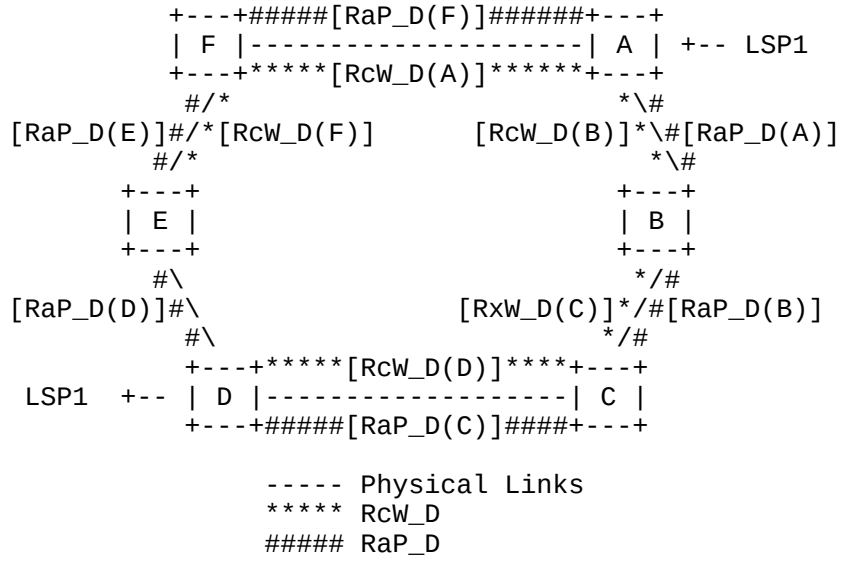

Figure 4: Label Operation of MSRP

4.2. Failure Detection

The MPLS-TP section-layer OAM is used to monitor the connectivity between each two adjacent nodes on the ring using the mechanisms defined in [RFC6371]. Protection switching is triggered by the failure detected on the ring by the OAM mechanisms.

Two ports of a link form a Maintenance Entity Group (MEG), and a MEG End Point (MEP) function is installed in each ring port. Continuity Check (CC) OAM packets are periodically exchanged between each pair of MEPs to monitor the link health. Three consecutive lost CC packets MUST be interpreted as a link failure.

A node failure is regarded as the failure of two links attached to that node. The two nodes adjacent to the failed node detect the failure in the links that are connected to the failed node.

4.3. Ring Protection

This section specifies the ring protection mechanisms in detail. In general, the description uses the clockwise working ring tunnel and the corresponding anticlockwise protection ring tunnel as an example, but the mechanism is applicable in the same way to the anticlockwise working and clockwise protection ring tunnels.

In a ring network, each working ring tunnel is associated with a protection ring tunnel in the opposite direction, and every node MUST obtain the ring topology either by configuration or via a topology discovery mechanism. The ring topology and the connectivity (Intact or Severed) between two adjacent ring nodes form the ring map. Each ring node maintains the ring map and uses it to perform ring protection switching.

Taking the topology in Figure 4 as an example, LSP1 enters the ring at Node A and leaves the ring at Node D. In normal state, LSP1 is carried by the clockwise working ring tunnel (RcW_D) through the path A->B->C->D. The label operation is:

[LSP1](Payload) -> [RCW_D(B)|LSP1](NodeA) -> [RCW_D(C)|LSP1](NodeB) -> [RCW_D(D)| LSP1](NodeC) -> [LSP1](Payload).

Then at Node D, the packet will be forwarded based on the label stack of LSP1.

Three typical ring protection mechanisms are described in this section: wrapping, short-wrapping, and steering. All nodes on the same ring MUST use the same protection mechanism. If the RPS protocol in any node detects an RPS message with a protection-switching mode that was not provisioned in that node, a failure of protocol will be reported, and the protection mechanism will not be activated.

Wrapping ring protection: the node that detects a failure or accepts a switch request switches the traffic impacted by the failure or the switch request to the opposite direction (away from the failure). In

this way, the impacted traffic is switched to the protection ring tunnel by the switching node upstream of the failure, then it travels around the ring to the switching node downstream of the failure through the protection ring tunnel, where it is switched back onto the working ring tunnel to reach the egress node.

Short-wrapping ring protection provides some optimization to wrapping protection, in which the impacted traffic is only switched once to the protection ring tunnel by the switching node upstream to the failure. At the egress node, the traffic leaves the ring from the protection ring tunnel. This can reduce the traffic detour of wrapping protection.

Steering ring protection implies that the node that detects a failure sends a request along the ring to the other node adjacent to the failure, and all nodes in the ring process this information. For the impacted traffic, the ingress node (which adds traffic to the ring) performs switching of the traffic from working to the protection ring tunnel, and the egress node will drop the traffic received from the protection ring tunnel.

The following sections describe these protection mechanisms in detail.

4.3.1. Wrapping

With the wrapping mechanism, the protection ring tunnel is a closed ring identified by the egress node. As shown in Figure 4, the RaP_D is the anticlockwise protection ring tunnel for the clockwise working ring tunnel RcW_D. As specified in the following sections, the closed ring protection tunnel can protect both link failures and node failures. Wrapping can be applicable for the protection of Point-to-Multipoint (P2MP) LSPs on the ring; the details of which are outside the scope of this document.

4.3.1.1. Wrapping for Link Failure

When a link failure between Nodes B and C occurs, if it is a bidirectional failure, both Nodes B and C can detect the failure via the OAM mechanism; if it is a unidirectional failure, one of the two nodes would detect the failure via the OAM mechanism. In both cases, the node at the other side of the detected failure will be determined by the ring map and informed using the RPS protocol, which is specified in Section 5. Then Node B switches the clockwise working ring tunnel (RcW_D) to the anticlockwise protection ring tunnel (RaP_D), and Node C switches the anticlockwise protection ring tunnel (RaP_D) back to the clockwise working ring tunnel (RcW_D). The

payload that enters the ring at Node A and leaves the ring at Node D follows the path A->B->A->F->E->D->C->D. The label operation is:

[LSP1](Payload) -> [RcW_D(B)|LSP1](Node A) -> [RaP_D(A)|LSP1](Node B) -> [RaP_D(F)|LSP1](Node A) -> [RaP_D(E)|LSP1] (Node F) -> [RaP_D(D)|LSP1] (Node E) -> [RaP_D(C)|LSP1] (Node D) -> [RcW_D(D)|LSP1](Node C) -> [LSP1](Payload).

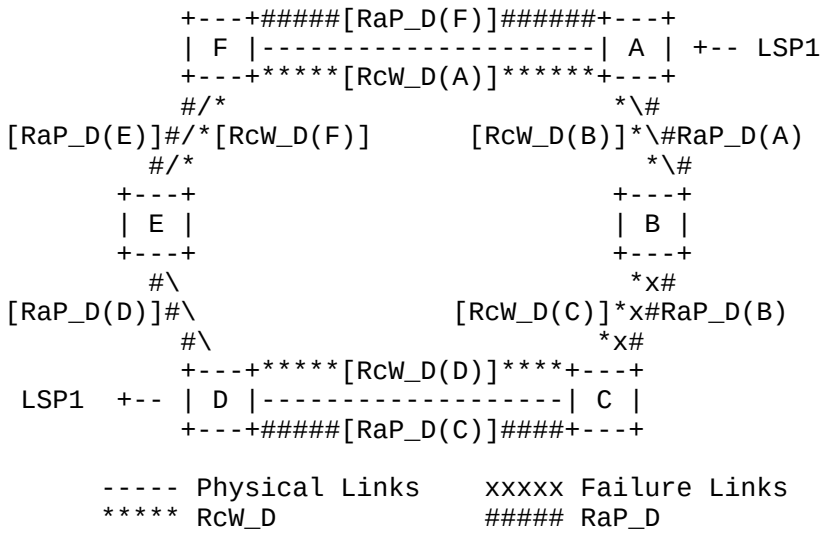

Figure 5: Wrapping for Link Failure

4.3.1.2. Wrapping for Node Failure

As shown in Figure 6, when Node B fails, Node A detects the failure between A and B and switches the clockwise working ring tunnel (RcW_D) to the anticlockwise protection ring tunnel (RaP_D); Node C detects the failure between C and B and switches the anticlockwise protection ring tunnel (RaP_D) to the clockwise working ring tunnel (RcW_D). The node at the other side of the failed node will be determined by the ring map and informed using the RPS protocol specified in Section 5.

The payload that enters the ring at Node A and exits at Node D follows the path A->F->E->D->C->D. The label operation is:

[LSP1](Payload)-> [RaP_D(F)|LSP1](NodeA) -> [RaP_D(E)|LSP1](NodeF) -> [RaP_D(D)|LSP1](NodeE) -> [RaP_D(C)|LSP1] (NodeD) -> [RcW_D(D)|LSP1] (NodeC) -> [LSP1](Payload).

In one special case where Node D fails, all the ring tunnels with Node D as the egress will become unusable. The ingress node will update its ring map according to received RPS messages and determine that the egress node is not reachable; thus, it will not send traffic to either the working or the protection tunnel. However, before the failure location information is propagated to all the ring nodes, the wrapping protection mechanism may cause a temporary traffic loop: Node C detects the failure and switches the traffic from the clockwise working ring tunnel (RcW_D) to the anticlockwise protection ring tunnel (RaP_D); Node E also detects the failure and switches the traffic from the anticlockwise protection ring tunnel (RaP_D) back to the clockwise working ring tunnel (RcW_D). A possible mechanism to mitigate the temporary loop problem is: the TTL of the ring tunnel label is set to 2*N by the ingress ring node of the traffic, where N is the number of nodes on the ring.

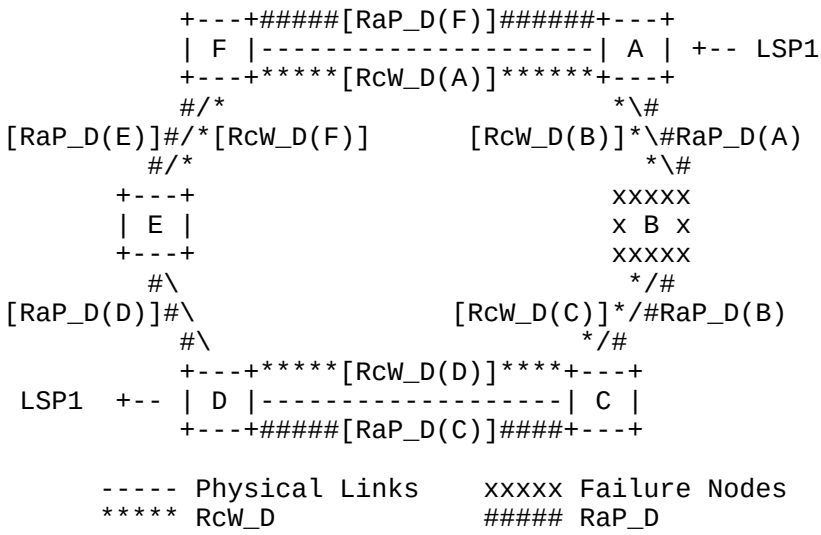

Figure 6: Wrapping for Node Failure

4.3.2. Short-Wrapping

With the wrapping protection scheme, protection switching is executed at both nodes adjacent to the failure; consequently, the traffic will be wrapped twice. This mechanism will cause additional latency and bandwidth consumption when traffic is switched to the protection path.

With short-wrapping protection, protection switching is executed only at the node upstream to the failure, and the packet leaves the ring in the protection ring tunnel at the egress node. This scheme can reduce the additional latency and bandwidth consumption when traffic is switched to the protection path. However, the two directions of a protected bidirectional LSP are no longer co-routed under the protection-switching conditions.

In the traditional wrapping solution, the protection ring tunnel is configured as a closed ring, while in the short-wrapping solution, the protection ring tunnel is configured as ended at the egress node, which is similar to the working ring tunnel. Short-wrapping is easy to implement in shared-ring protection because both the working and protection ring tunnels are terminated on the egress nodes. Figure 7 shows the clockwise working ring tunnel and the anticlockwise protection ring tunnel with Node D as the egress node.

4.3.2.1. Short-Wrapping for Link Failure

As shown in Figure 7, in normal state, LSP1 is carried by the clockwise working ring tunnel (RcW_D) through the path A->B->C->D. When a link failure between Nodes B and C occurs, Node B switches the working ring tunnel RcW_D to the protection ring tunnel RaP_D in the opposite direction. The difference with wrapping occurs in the protection ring tunnel at the egress node. In short-wrapping protection, Rap_D ends in Node D, and then traffic will be forwarded based on the LSP labels. Thus, with the short-wrapping mechanism, LSP1 will follow the path A->B->A->F->E->D when a link failure between Node B and Node C happens. The protection switch at Node D is based on the information from its ring map and the information received via the RPS protocol.

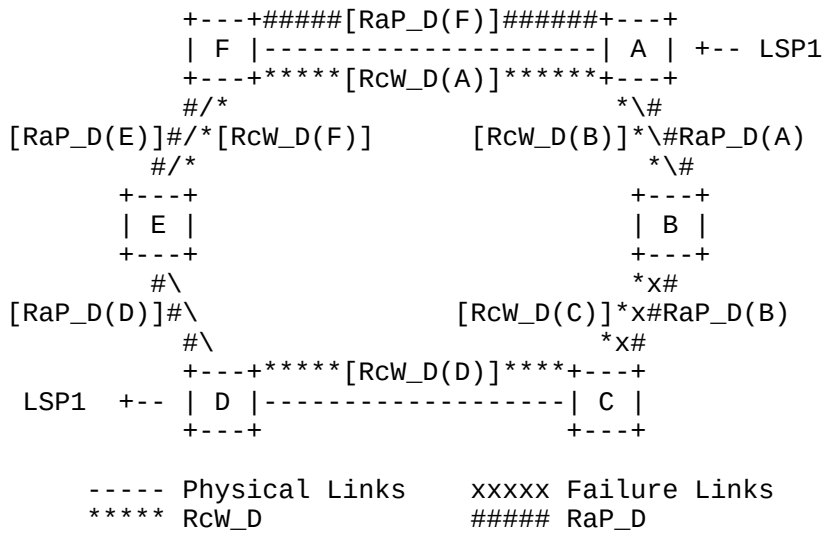

Figure 7: Short-Wrapping for Link Failure

4.3.2.2. Short-Wrapping for Node Failure

For the node failure that happens on a non-egress node, the short-wrapping protection switching is similar to the link failure case as described in the previous section. This section specifies the scenario of an egress node failure.

As shown in Figure 8, LSP1 enters the ring on Node A and leaves the ring on Node D. In normal state, LSP1 is carried by the clockwise working ring tunnel (RcW_D) through the path A->B->C->D. When Node D fails, the traffic of LSP1 cannot be protected by any ring tunnels that use Node D as the egress node. The ingress node will update its ring map according to received RPS messages and determine that the egress node is not reachable; thus, it will not send traffic to either the working or the protection tunnel. However, before the failure location information is propagated to all the ring nodes using the RPS protocol, Node C switches all the traffic on the working ring tunnel RcW_D to the protection ring tunnel RaP_D in the opposite direction based on the information in the ring map. When the traffic arrives at Node E, which also detects the failure of Node D, the protection ring tunnel RaP_D cannot be used to forward traffic to Node D. With the short-wrapping mechanism, protection switching can only be performed once from the working ring tunnel to the protection ring tunnel; thus, Node E MUST NOT switch the traffic that is already carried on the protection ring tunnel back to the working

ring tunnel in the opposite direction. Instead, Node E will discard the traffic received on RaP_D locally. This can avoid the temporary traffic loop when the failure happens on the egress node of the ring tunnel. This also illustrates one of the benefits of having separate working and protection ring tunnels in each ring direction.

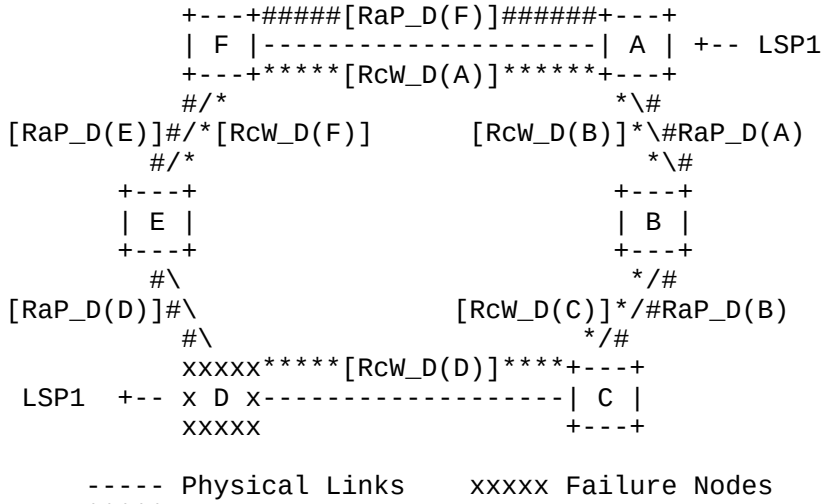

Figure 8: Short-Wrapping for Egress Node Failure

4.3.3. Steering

With the steering protection mechanism, the ingress node (which adds traffic to the ring) performs switching from the working to the protection ring tunnel, and at the egress node, the traffic leaves the ring from the protection ring tunnel.

When a failure occurs in the ring, the node that detects the failure with an OAM mechanism sends the failure information in the opposite direction of the failure hop by hop along the ring using an RPS request message and the ring-map information. When a ring node receives the RPS message that identifies a failure, it can determine the location of the fault by using the topology information of the ring map and updating the ring map accordingly; then, it can determine whether the LSPs entering the ring locally need to switch over or not. For LSPs that need to switch over, it will switch the LSPs from the working ring tunnels to their corresponding protection ring tunnels.

4.3.3.1. Steering for Link Failure

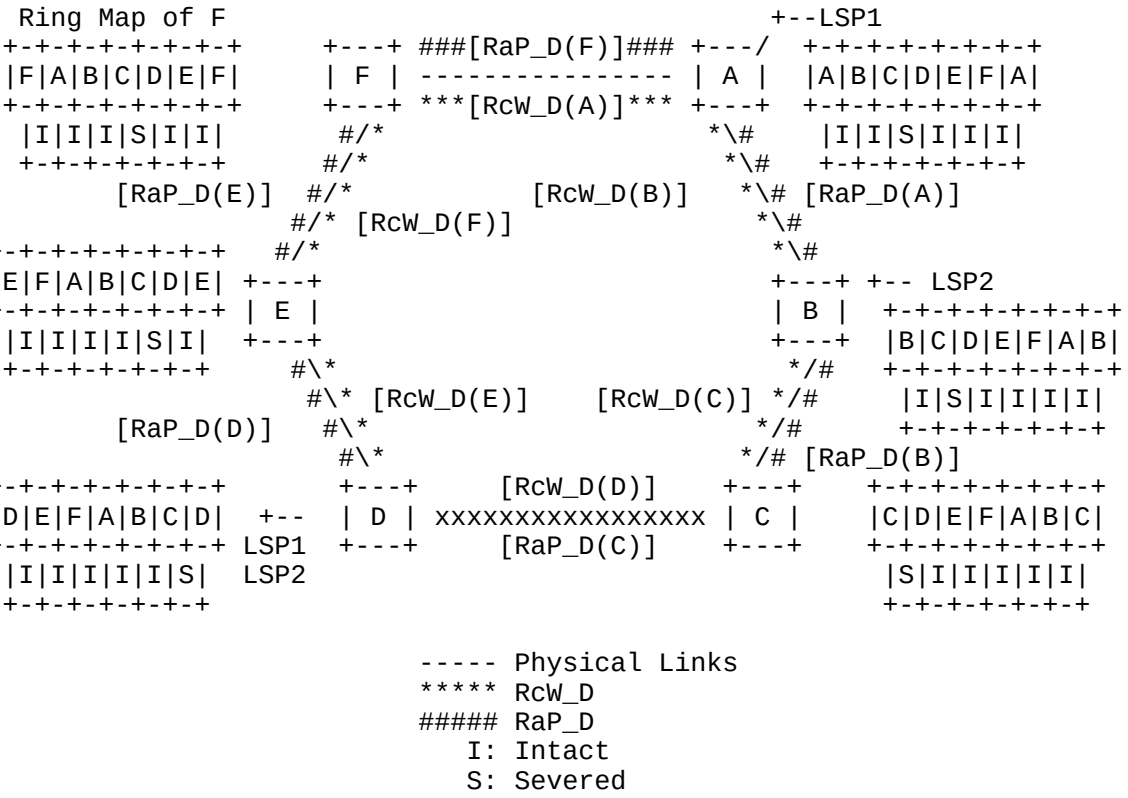

Figure 9: Steering Operation and Protection Switching When Link C-D Fails

As shown in Figure 9, LSP1 enters the ring from Node A while LSP2 enters the ring from Node B, and both of them have the same destination, which is Node D.

In normal state, LSP1 is carried by the clockwise working ring tunnel (RcW_D) through the path A->B->C->D, and the label operation is: [LSP1](Payload) -> [RcW_D(B)|LSP1](NodeA) -> [RcW_D(C)| LSP1](NodeB) -> [RcW_D(D)|LSP1](NodeC) -> [LSP1](Payload).

LSP2 is carried by the clockwise working ring tunnel (RcW_D) through the path B->C->D, and the label operation is: [LSP2](Payload) -> [RcW_D(C)|LSP2](NodeB) -> [RcW_D(D)|LSP2](NodeC) -> [LSP2](Payload).

If the link between Nodes C and D fails, according to the fault detection and distribution mechanisms, Node D will find out that there is a failure in the link between C and D, and it will update the link state of its ring topology, changing the link between C and D from normal to fault. In the direction that is opposite to the failure position, Node D will send the state report message to Node E, informing Node E of the fault between C and D, and E will update the link state of its ring topology accordingly, changing the link between C and D from normal to fault. In this way, the state report message is sent hop by hop in the clockwise direction. Similar to Node D, Node C will send the failure information in the anticlockwise direction.

When Node A receives the failure report message and updates the link state of its ring map, it is aware that there is a fault on the clockwise working ring tunnel to Node D (RcW_D), and LSP1 enters the ring locally and is carried by this ring tunnel; thus, Node A will decide to switch the LSP1 onto the anticlockwise protection ring tunnel to Node D (RaP_D). After the switchover, LSP1 will follow the path A->F->E->D, and the label operation is: [LSP1](Payload) -> [RaP_D(F)| LSP1](NodeA) -> [RaP_D(E)|LSP1](NodeF) -> [RaP_D(D)|LSP1](NodeE) -> [LSP1](Payload).

The same procedure also applies to the operation of LSP2. When Node B updates the link state of its ring topology, and finds out that the working ring tunnel RcW_D has failed, it will switch the LSP2 to the anticlockwise protection tunnel RaP_D. After the switchover, LSP2 goes through the path B->A->F->E->D, and the label operation is: [LSP2](Payload) -> [RaP_D(A)|LSP2](NodeB) -> [RaP_D(F)|LSP2](NodeA) -> [RaP_D(E)|LSP2](NodeF) -> [RaP_D(D)|LSP2](NodeE) -> [LSP2](Payload).

Assume the link between Nodes A and B breaks down, as shown in Figure 10. Similar to the above failure case, Node B will detect a fault in the link between A and B, and it will update its ring map, changing the link state between A and B from normal to fault. The state report message is sent hop by hop in the clockwise direction, notifying every node that there is a fault between Nodes A and B, and every node updates the link state of its ring topology. As a result, Node A will detect a fault in the working ring tunnel to Node D, and switch LSP1 to the protection ring tunnel, while Node B determines that the working ring tunnel for LSP2 still works fine, and it will not perform the switchover.

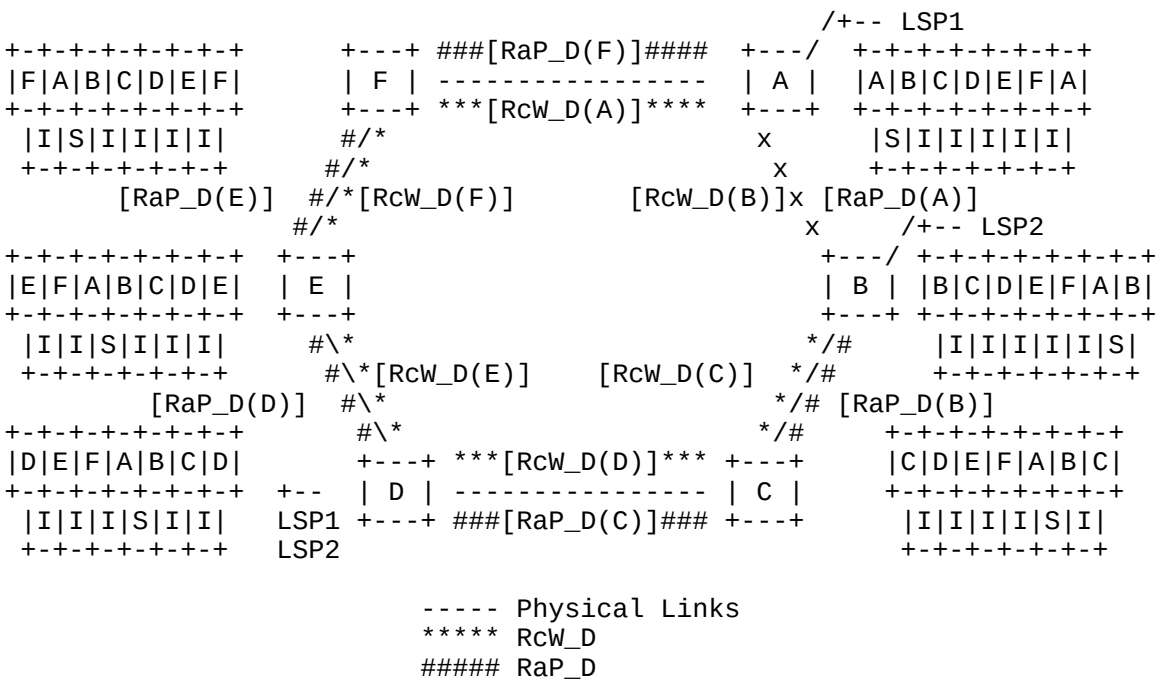

Figure 10: Steering Operation and Protection Switching When Link A-B Fails

4.3.3.2. Steering for Node Failure

For a node failure that happens on a non-egress node, steering protection switching is similar to the link failure case as described in the previous section.

If the failure occurs at the egress node of the LSP, the ingress node will update its ring map according to the received RPS messages; it will also determine that the egress node is not reachable after the failure, thus it will not send traffic to either the working or the protection tunnel, and a traffic loop can be avoided.

4.4. Interconnected Ring Protection

4.4.1. Interconnected Ring Topology

Interconnected ring topology is widely used in MPLS-TP networks. For a given ring, the interconnection node acts as the egress node for that ring, meaning that all LSPs using the interconnection node as an egress from one specific ring to another will use the same group of ring tunnels within the ring. This document will discuss two typical interconnected ring topologies:

1. Single-node interconnected rings

   In single-node interconnected rings, the connection between the two rings is through a single node. Because the interconnection node is in fact a single point of failure, this topology should be avoided in real transport networks.

   Figure 11 shows the topology of single-node interconnected rings. Node C is the interconnection node between Ring1 and Ring2.

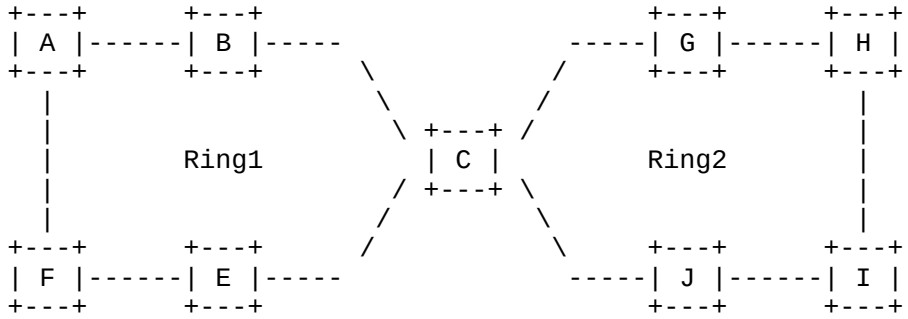

Figure 11: Single-Node Interconnected Rings

2. Dual-node interconnected rings

   In dual-node interconnected rings, the connection between the two rings is through two nodes. The two interconnection nodes belong to both interconnected rings. This topology can recover from one interconnection node failure.

   Figure 12 shows the topology of dual-node interconnected rings. Nodes C and D are the interconnection nodes between Ring1 and Ring2.

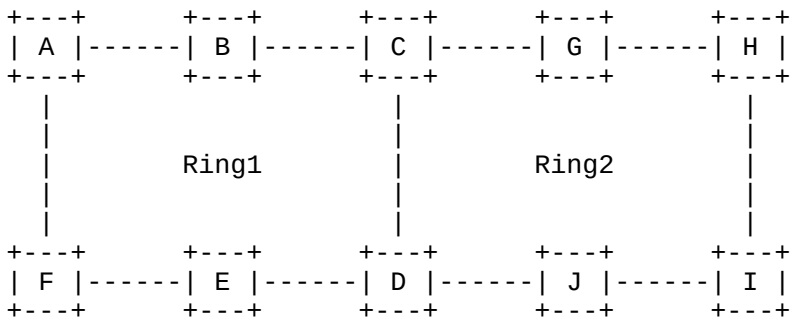

Figure 12: Dual-Node Interconnected Rings

4.4.2. Interconnected Ring Protection Mechanisms

Interconnected rings can be treated as two independent rings. The RPS protocol operates on each ring independently. A failure that happens in one ring only triggers protection switching in the ring itself and does not affect the other ring, unless the failure is on the interconnection node. In this way, protection switching on each ring is the same as the mechanisms described in Section 4.3.

The service LSPs that traverse the interconnected rings use the ring tunnels in each ring; within a given ring, the tunnel is selected using normal ring-selection procedures. The traversing LSPs are stitched on the interconnection node. On the interconnection node, the ring tunnel label of the source ring is popped, then LSP label is swapped; after that, the ring tunnel label of the destination ring is pushed.

In the dual-node interconnected ring scenario, the two interconnection nodes can be managed as a virtual node group. In addition to the ring tunnels to each physical ring node, each ring SHOULD assign the working and protection ring tunnels to the virtual interconnection node group. In addition, on both nodes in the virtual interconnection node group, the same LSP label is assigned for each traversed LSP. This way, any interconnection node in the virtual node group can terminate the working or protection ring tunnels targeted to the virtual node group and stitch the service LSP from the source ring tunnel to the destination ring tunnel.

When the service LSP passes through the interconnected rings, the direction of the working ring tunnels used on both rings SHOULD be the same. In dual-node interconnected rings, this ensures that in normal state the traffic passes only one of the two interconnection nodes and does not pass the link between the two interconnection

nodes. The traffic will then only be switched to the protection path if the interconnection node that is in working path fails. For example, if the service LSP uses the clockwise working ring tunnel on Ring1, when the service LSP leaves Ring1 and enters Ring2, the working ring tunnel used on Ring2 should also follow the clockwise direction.

4.4.3. Ring Tunnels in Interconnected Rings

The same ring tunnels as described in Section 4.1 are used in each ring of the interconnected rings. In addition, ring tunnels to the virtual interconnection node group are established on each ring of the interconnected rings, that is:

- one clockwise working ring tunnel to the virtual interconnection node group
- one anticlockwise protection ring tunnel to the virtual interconnection node group
- one anticlockwise working ring tunnel to the virtual interconnection node group
- one clockwise protection ring tunnel to the virtual interconnection node group

The ring tunnels to the virtual interconnection node group are shared by all LSPs that need to be forwarded to other rings. These ring tunnels can terminate at any node in the virtual interconnection node group.

For example, all the ring tunnels on Ring1 in Figure 13 are provisioned as follows:

- To Node A: R1cW_A, R1aW_A, R1cP_A, R1aP_A
- To Node B: R1cW_B, R1aW_B, R1cP_B, R1aP_B
- To Node C: R1cW_C, R1aW_C, R1cP_C, R1aP_C
- To Node D: R1cW_D, R1aW_D, R1cP_D, R1aP_D
- To Node E: R1cW_E, R1aW_E, R1cP_E, R1aP_E
- To Node F: R1cW_F, R1aW_F, R1cP_F, R1aP_F
- To the virtual interconnection node group (including Nodes F and A): R1cW_F&A, R1aW_F&A, R1cP_F&A, R1aP_F&A

All the ring tunnels on Ring2 in Figure 13 are provisioned as follows:

- To Node A: R2cW_A, R2aW_A, R2cP_A, R2aP_A
- To Node F: R2cW_F, R2aW_F, R2cP_F, R2aP_F
- To Node G: R2cW_G, R2aW_G, R2cP_G, R2aP_G
- To Node H: R2cW_H, R2aW_H, R2cP_H, R2aP_H
- To Node I: R2cW_I, R2aW_I, R2cP_I, R2aP_I
- To Node J: R2cW_J, R2aW_J, R2cP_J, R2aP_J
- To the virtual interconnection node group (including Nodes F and A): R2cW_F&A, R2aW_F&A, R2cP_F&A, R2aP_F&A

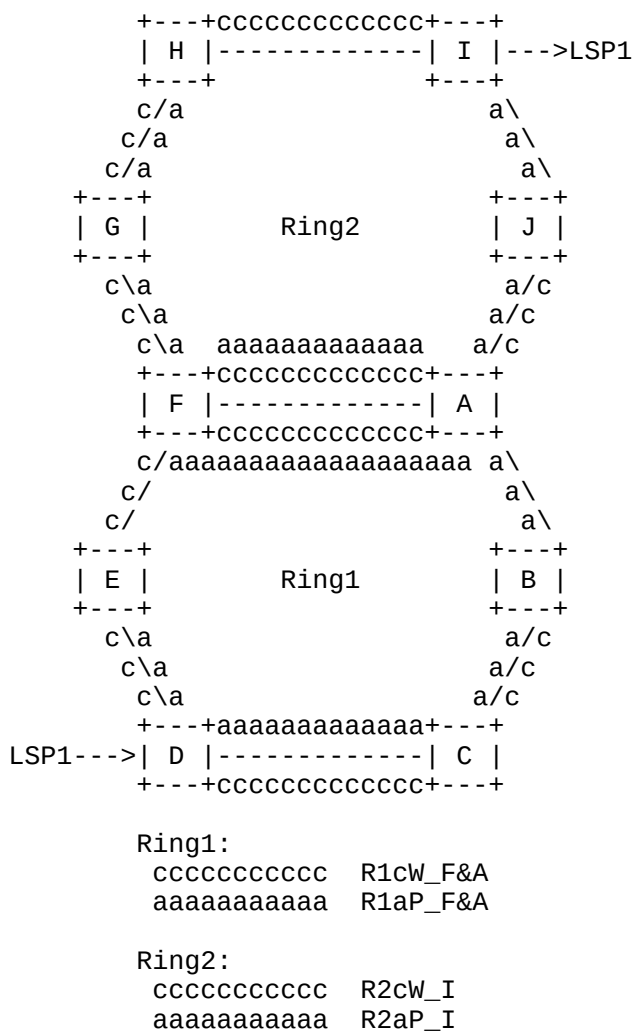

Figure 13: Ring Tunnels for the Interconnected Rings

4.4.4. Interconnected Ring-Switching Procedure

As shown in Figure 13, for the service LSP1 that enters Ring1 at Node D and leaves Ring1 at Node F and continues to enter Ring2 at Node F and leaves Ring2 at Node I, the short-wrapping protection scheme is described as below.

In normal state, LSP1 follows R1cW_F&A in Ring1 and R2cW_I in Ring2. At the interconnection Node F, the label used for the working ring tunnel R1cW_F&A in Ring1 is popped, the LSP label is swapped, and the label used for the working ring tunnel R2cW_I in Ring2 will be pushed based on the inner LSP label lookup. The working path that the service LSP1 follows is: LSP1->R1cW_F&A (D->E->F)->R2cW_I(F->G->H->I)->LSP1.

In case of link failure, for example, when a failure occurs on the link between Nodes F and E, Node E will detect the failure and execute protection switching as described in Section 4.3.2. The path that the service LSP1 follows after switching change to: LSP1->R1cW_F &A(D->E)->R1aP_F&A(E->D->C->B->A)->R2cW_I(A->F->G->H->I)->LSP1.

In case of a non-interconnection node failure, for example, when the failure occurs at Node E in Ring1, Node D will detect the failure and execute protection switching as described in Section 4.3.2. The path that the service LSP1 follows after switching becomes: LSP1->R1aP_F&A(D->C->B->A)->R2cW_I(A->F->G->H->I)->LSP1.

In case of an interconnection node failure, for example, when the failure occurs at the interconnection Node F, Node E in Ring1 will detect the failure and execute protection switching as described in Section 4.3.2. Node A in Ring2 will also detect the failure and execute protection switching as described in Section 4.3.2. The path that the service traffic LSP1 follows after switching is: LSP1->R1cW_F&A(D->E)->R1aP_F&A(E->D->C->B->A)->R2aP_I(A->J->I)->LSP1.

4.4.5. Interconnected Ring Detection Mechanism

As shown in Figure 13, in normal state, the service traffic LSP1 traverses D->E->F in Ring1 and F->G->H->I in Ring2. Nodes A and F are the interconnection nodes. When both links between Nodes F and G and between Nodes F and A fail, the ring tunnel from Node F to Node I in Ring2 becomes unreachable. However, the other interconnection Node A is still available, and LSP1 can still reach Node I via Node A.

In order to achieve this, the interconnection nodes need to know the ring topology of each ring so that they can judge whether a node is reachable. This judgment is based on the knowledge of the ring map and the fault location. The ring map can be obtained from the Network Management System (NMS) or topology discovery mechanisms. The fault location can be obtained by transmitting the fault information around the ring. The nodes that detect the failure will transmit the fault information in the opposite direction hop by hop using the RPS protocol message. When the interconnection node receives the message that informs the failure, it will calculate the

location of the fault according to the topology information that is maintained by itself and determines whether the LSPs entering the ring at itself can reach the destination. If the destination node is reachable, the LSP will leave the source ring and enter the destination ring. If the destination node is not reachable, the LSP will switch to the anticlockwise protection ring tunnel.

In Figure 13, Node F determines that the ring tunnel to Node I is unreachable; the service LSP1 for which the destination node on Ring2 is Node I MUST switch to the protection ring tunnel (R1aP_F&A), and consequently, the service traffic LSP1 traverses the interconnected rings at Node A. Node A will pop the ring tunnel label of Ring1 and push the ring tunnel label of Ring2 and send the traffic to Node I via the ring tunnel (R2aW_I).

5. Ring Protection Coordination Protocol

5.1. RPS and PSC Comparison on Ring Topology

This section provides comparison between RPS and Protection State Coordination (PSC) [RFC6378] [RFC6974] on ring topologies. This can be helpful to explain the reason of defining a new protocol for ring protection switching.

The PSC protocol [RFC6378] is designed for point-to-point LSPs, on which the protection switching can only be performed on one or both of the endpoints of the LSP. The RPS protocol is designed for ring tunnels, which consist of multiple ring nodes, and the failure could happen on any segment of the ring; thus, RPS is capable of identifying and handling the different failures on the ring and coordinating the protection-switching behavior of all the nodes on the ring. As will be specified in the following sections, this is achieved with the introduction of the "pass-through" state for the ring nodes, and the location of the protection request is identified via the node IDs in the RPS request message.

Taking a ring topology with N nodes as an example:

With the mechanism specified in [RFC6974], on every ring node, a linear protection configuration has to be provisioned with every other node in the ring, i.e., with (N-1) other nodes. This means that on every ring node there will be (N-1) instances of the PSC protocol. And in order to detect faults and to transport the PSC message, each instance shall have a MEP on the working path and a MEP on the protection path, respectively. This means that every node on the ring needs to be configured with (N-1) * 2 MEPs.

With the mechanism defined in this document, on every ring node there will only be a single instance of the RPS protocol. In order to detect faults and to transport the RPS message, each node only needs to have a MEP on the section to its adjacent nodes, respectively. In this way, every ring node only needs to be configured with 2 MEPs.

As shown in the above example, RPS is designed for ring topologies and can achieve ring protection efficiently with minimum protection instances and OAM entities, which meets the requirements on topology-specific recovery mechanisms as specified in [RFC5654].

5.2. RPS Protocol

The RPS protocol defined in this section is used to coordinate the protection-switching action of all the ring nodes in the same ring.

The protection operation of the ring tunnels is controlled with the help of the RPS protocol. The RPS processes in each of the individual ring nodes that form the ring MUST communicate using the Generic Associated Channel (G-ACh). The RPS protocol is applicable to all the three ring protection modes. This section takes the short-wrapping mechanism described in Section 4.3.2 as an example.

The RPS protocol is used to distribute the ring status information and RPS requests to all the ring nodes. Changes in the ring status information and RPS requests can be initiated automatically based on link status or caused by external commands.

Each node on the ring is uniquely identified by assigning it a node ID. The node ID MUST be unique on each ring. The maximum number of nodes on the ring supported by the RPS protocol is 127. The node ID SHOULD be independent of the order in which the nodes appear on the ring. The node ID is used to identify the source and destination nodes of each RPS request.

Every node obtains the ring topology either by configuration or via some topology discovery mechanism. The ring map consists of the ring topology information, and connectivity status (Intact or Severed) between the adjacent ring nodes, which is determined via the OAM message exchanged between the adjacent nodes. The ring map is used by every ring node to determine the switchover behavior of the ring tunnels.

As shown in Figure 14, when no protection switching is active on the ring, each node MUST send RPS requests with No Request (NR) to its two adjacent nodes periodically. The transmission interval of RPS requests is specified in Section 5.2.1.

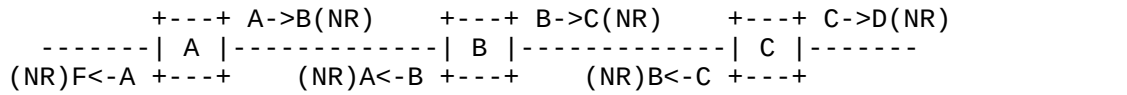

Figure 14: RPS Communication between the Ring Nodes in Case of No Failure in the Ring

As shown in Figure 15, when a node detects a failure and determines that protection switching is required, it MUST send the appropriate RPS request in both directions to the destination node. The destination node is the other node that is adjacent to the identified failure. When a node that is not the destination node receives an RPS request and it has no higher-priority local request, it MUST transfer in the same direction the RPS request as received. In this way, the switching nodes can maintain RPS protocol communication in the ring. The RPS request MUST be terminated by the destination node of the message. If an RPS request with the node itself set as the source node is received, this message MUST be dropped and not be forwarded to the next node.

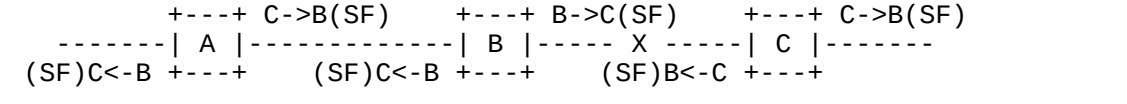

Figure 15: RPS Communication between the Ring Nodes in Case of Failure between Nodes B and C

Note that in the case of a bidirectional failure such as a cable cut, the two adjacent nodes detect the failure and send each other an RPS request in opposite directions.

- In rings utilizing the wrapping protection, each node detects the failure or receives the RPS request as the destination node MUST perform the switch from/to the working ring tunnels to/from the protection ring tunnels if it has no higher-priority active RPS request.

- In rings utilizing the short-wrapping protection, each node detects the failure or receives the RPS request as the destination node MUST perform the switch only from the working ring tunnels to the protection ring tunnels.

- In rings utilizing the steering protection, when a ring switch is required, any node MUST perform the switches if its added/dropped traffic is affected by the failure. Determination of the affected traffic MUST be performed by examining the RPS requests (indicating the nodes adjacent to the failure or failures) and the stored ring map (indicating the relative position of the failure and the added traffic destined towards that failure).

When the failure has cleared and the Wait-to-Restore (WTR) timer has expired, the nodes that generate the RPS requests MUST drop their respective switches and MUST generate an RPS request carrying the NR code. The node receiving such an RPS request from both directions MUST drop its protection switches.

A protection switch MUST be initiated by one of the criteria specified in Section 5.3. A failure of the RPS protocol or controller MUST NOT trigger a protection switch.

Ring switches MUST be preempted by higher-priority RPS requests. For example, consider a protection switch that is active due to a manual switch request on the given link, and another protection switch is required due to a failure on another link. Then an RPS request MUST be generated, the former protection switch MUST be dropped, and the latter protection switch established.

The MPLS-TP Shared-Ring Protection mechanism supports multiple protection switches in the ring, resulting in the ring being segmented into two or more separate segments. This may happen when several RPS requests of the same priority exist in the ring due to multiple failures or external switch commands.

Proper operation of the MSRP mechanism relies on all nodes using their ring map to determine the state of the ring (nodes and links). In order to accommodate ring state knowledge, the RPS requests MUST be sent in both directions during a protection switch.

5.2.1. Transmission and Acceptance of RPS Requests

A new RPS request MUST be transmitted immediately when a change in the transmitted status occurs.

The first three RPS protocol messages carrying a new RPS request MUST be transmitted as fast as possible. For fast protection switching within 50 ms, the interval of the first three RPS protocol messages SHOULD be 3.3 ms. The successive RPS requests SHOULD be transmitted with the interval of 5 seconds. A ring node that is not the destination of the received RPS message MUST forward it to the next node along the ring immediately.

5.2.2. RPS Protocol Data Unit (PDU) Format

Figure 16 depicts the format of an RPS packet that is sent on the G-ACh. The Channel Type field is set to indicate that the message is an RPS message.

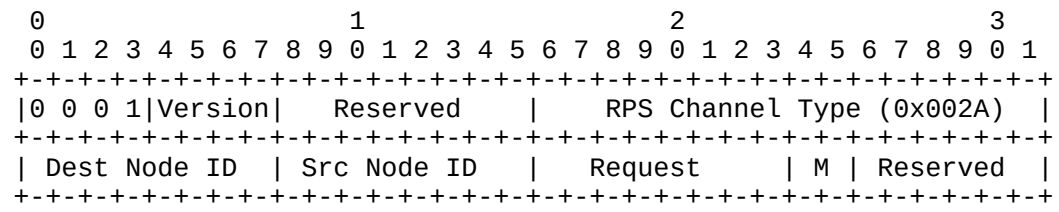

Figure 16: G-ACh RPS Packet Format

The following fields MUST be provided:

- Destination Node ID: The destination node ID MUST always be set to the value of the node ID of the adjacent node. The node ID MUST be unique on each ring. Valid destination node ID values are 1-127.

- Source Node ID: The source node ID MUST always be set to the ID value of the node generating the RPS request. The node ID MUST be unique on each ring. Valid source node ID values are 1-127.

- Protection-Switching Mode (M): This 2-bit field indicates the protection-switching mode used by the sending node of the RPS message. This can be used to check that the ring nodes on the same ring use the same protection-switching mechanism. The defined values of the M field are listed as below:

Bits (MSB - LSB)	Protection-Switching Mode
0 0	Reserved
0 1	Wrapping
1 0	Short-Wrapping
1 1	Steering

Note:
MSB = most significant bit
LSB = least significant bit

- RPS Request Code: A code consisting of 8 bits as specified below:

Bits (MSB - LSB)	Condition, State, or External Request	Priority
0 0 0 0 1 1 1 1	Lockout of Protection (LP)	highest
0 0 0 0 1 1 0 1	Forced Switch (FS)	
0 0 0 0 1 0 1 1	Signal Fail (SF)	
0 0 0 0 0 1 1 0	Manual Switch (MS)	
0 0 0 0 0 1 0 1	Wait-to-Restore (WTR)	
0 0 0 0 0 0 1 1	Exercise (EXER)	
0 0 0 0 0 0 0 1	Reverse Request (RR)	
0 0 0 0 0 0 0 0	No Request (NR)	lowest

5.2.3. Ring Node RPS States

Idle state: A node is in the idle state when it has no RPS request and is sending and receiving an NR code to/from both directions.

Switching state: A node not in the idle or pass-through states is in the switching state.

Pass-through state: A node is in the pass-through state when its highest priority RPS request is a request not destined to it or generated by it. The pass-through is bidirectional.

5.2.3.1. Idle State

A node in the idle state MUST generate the NR request in both directions.

A node in the idle state MUST terminate RPS requests that flow in both directions.

A node in the idle state MUST block the traffic flow on protection ring tunnels in both directions.

5.2.3.2. Switching State

A node in the switching state MUST generate an RPS request to its adjacent node with its highest RPS request code in both directions when it detects a failure or receives an external command.

In a bidirectional failure condition, both of the nodes adjacent to the failure detect the failure and send the RPS request in both directions with the destination set to each other; while each node can only receive the RPS request via the long path, the message sent via the short path will get lost due to the bidirectional failure. Here, the short path refers to the shorter path on the ring between the source and destination node of the RPS request, and the long path refers to the longer path on the ring between the source and destination node of the RPS request. Upon receipt of the RPS request on the long path, the destination node of the RPS request MUST send an RPS request with its highest request code periodically along the long path to the other node adjacent to the failure.

In a unidirectional failure condition, the node that detects the failure MUST send the RPS request in both directions with the destination node set to the other node adjacent to the failure. The destination node of the RPS request cannot detect the failure itself but will receive an RPS request from both the short path and the long path. The destination node MUST acknowledge the received RPS requests by replying with an RPS request with the RR code on the short path and an RPS request with the received RPS request code on the long path. Accordingly, when the node that detects the failure receives the RPS request with RR code on the short path, then the RPS request received from the same node along the long path SHOULD be ignored.

A node in the switching state MUST terminate the received RPS requests in both directions and not forward it further along the ring.

The following switches as defined in Section 5.3.1 MUST be allowed to coexist:

- LP and LP
- FS and FS
- SF and SF
- FS and SF

When multiple MS RPS requests exist at the same time addressing different links and there is no higher-priority request on the ring, no switch SHOULD be executed and existing switches MUST be dropped. The nodes MUST still signal an RPS request with the MS code.

Multiple EXER requests MUST be allowed to coexist in the ring.

A node in a ring-switching state that receives the external command LP for the affected link MUST drop its switch and MUST signal NR for the locked link if there is no other RPS request on another link. The node still SHOULD signal a relevant RPS request for another link.

5.2.3.3. Pass-Through State

When a node is in a pass-through state, it MUST transfer the received RPS request unchanged in the same direction.

When a node is in a pass-through state, it MUST enable the traffic flow on protection ring tunnels in both directions.

5.2.4. RPS State Transitions

All state transitions are triggered by an incoming RPS request change, a WTR expiration, an externally initiated command, or locally detected MPLS-TP section failure conditions.

RPS requests due to a locally detected failure, an externally initiated command, or a received RPS request shall preempt existing RPS requests in the prioritized order given in Section 5.2.2, unless the requests are allowed to coexist.

5.2.4.1. Transitions between Idle and Pass-Through States

The transition from the idle state to pass-through state MUST be triggered by a valid RPS request change, in any direction, from the NR code to any other code, as long as the new request is not destined to the node itself. Both directions move then into a pass-through state, so that traffic entering the node through the protection ring tunnels are transferred transparently through the node.

A node MUST revert from pass-through state to the idle state when an RPS request with an NR code is received in both directions. Then both directions revert simultaneously from the pass-through state to the idle state.

5.2.4.2. Transitions between Idle and Switching States

Transition of a node from the idle state to the switching state MUST be triggered by one of the following conditions:

- A valid RPS request change from the NR code to any code received on either the long or the short path and is destined to this node
- An externally initiated command for this node

- The detection of an MPLS-TP section-layer failure at this node

Actions taken at a node in the idle state upon transition to the switching state are:

- For all protection-switch requests, except EXER and LP, the node MUST execute the switch
- For EXER, and LP, the node MUST signal the appropriate request but not execute the switch

In one of the following conditions, transition from the switching state to the idle state MUST be triggered:

- On the node that triggers the protection switching, when the WTR time expires or an externally initiated command is cleared, the node MUST transit from switching state to Idle State and signal the NR code using RPS message in both directions.
- On the node that enters the switching state due to the received RPS request: upon reception of the NR code from both directions, the head-end node MUST drop its switch, transition to idle state, and signal the NR code in both directions.

5.2.4.3. Transitions between Switching States

When a node that is currently executing any protection switch receives a higher-priority RPS request (due to a locally detected failure, an externally initiated command, or a ring protection switch request destined to it) for the same link, it MUST update the priority of the switch it is executing to the priority of the received RPS request.

When a failure condition clears at a node, the node MUST enter WTR condition and remain in it for the appropriate time-out interval, unless:

- A different RPS request with a higher priority than WTR is received
- Another failure is detected
- An externally initiated command becomes active

The node MUST send out a WTR code on both the long and short paths.

When a node that is executing a switch in response to an incoming SF RPS request (not due to a locally detected failure) receives a WTR code (unidirectional failure case), it MUST send out the RR code on the short path and the WTR on the long path.

5.2.4.4. Transitions between Switching and Pass-Through States

When a node that is currently executing a switch receives an RPS request for a non-adjacent link of higher priority than the switch it is executing, it MUST drop its switch immediately and enter the pass-through state.

The transition of a node from pass-through to switching state MUST be triggered by:

- An equal priority, a higher priority, or an allowed coexisting externally initiated command
- The detection of an equal priority, a higher priority, or an allowed coexisting automatic initiated command
- The receipt of an equal, a higher priority, or an allowed coexisting RPS request destined to this node

5.3. RPS State Machine

5.3.1. Switch Initiation Criteria

5.3.1.1. Administrative Commands

Administrative commands can be initiated by the network operator through the Network Management System (NMS). The operator command may be transmitted to the appropriate node via the MPLS-TP RPS message.

The following commands can be transferred by the RPS message:

- Lockout of Protection (LP): This command prevents any protection activity and prevents using ring switches anywhere in the ring. If any ring switches exist in the ring, this command causes the switches to drop.

- Forced Switch (FS) to protection: This command performs the ring switch of normal traffic from the working entity to the protection entity for the link between the node at which the command is initiated and the adjacent node to which the command is directed. This switch occurs regardless of the state of the MPLS-TP section for the requested link, unless a higher-priority switch request exists.

- Manual Switch (MS) to protection: This command performs the ring switch of the normal traffic from the working entity to the protection entity for the link between the node at which the command is initiated and the adjacent node to which the command is directed. This occurs if the MPLS-TP section for the requested link is not satisfying an equal or higher priority switch request.

- Exercise (EXER): This command exercises ring protection switching on the addressed link without completing the actual switch. The command is issued and the responses (RRs) are checked, but no normal traffic is affected.

The following commands are not transferred by the RPS message:

- Clear: This command clears the administrative command and WTR timer at the node to which the command was addressed. The node-to-node signaling after the removal of the externally initiated commands is performed using the NR code.

- Lockout of Working (LW): This command prevents the normal traffic transported over the addressed link from being switched to the protection entity by disabling the node’s capability of requesting a switch for this link in case of failure. If any normal traffic is already switched on the protection entity, the switch is dropped. If no other switch requests are active on the ring, the NR code is transmitted. This command has no impact on any other link. If the node receives the switch request from the adjacent node from any side, it will perform the requested switch. If the node receives the switch request addressed to the other node, it will enter the pass-through state.

5.3.1.2. Automatically Initiated Commands

Automatically initiated commands can be initiated based on MPLS-TP section-layer OAM indication and the received switch requests.

The node can initiate the following switch requests automatically:

- Signal Fail (SF): This command is issued when the MPLS-TP section-layer OAM detects a signal failure condition.

- Wait-to-Restore (WTR): This command is issued when the MPLS-TP section detects that the SF condition has cleared. It is used to maintain the state during the WTR period unless it is preempted by a higher-priority switch request. The WTR time may be configured by the operator in 1 minute steps between 0 and 12 minutes; the default value is 5 minutes.

- Reverse Request (RR): This command is transmitted to the source node of the received RPS message over the short path as an acknowledgment for receiving the switch request.

5.3.2. Initial States

This section describes the possible states of a ring node, the corresponding action of the working and protection ring tunnels on the node, and the RPS request that should be generated in that state.

	State	Signaled RPS
A	Idle Working: no switch Protection: no switch	NR
B	Pass-through Working: no switch Protection: pass-through	N/A
C	Switching - LP Working: no switch Protection: no switch	LP
D	Idle - LW Working: no switch Protection: no switch	NR
E	Switching - FS Working: switched Protection: switched	FS
F	Switching - SF Working: switched Protection: switched	SF
G	Switching - MS Working: switched Protection: switched	MS
H	Switching - WTR Working: switched Protection: switched	WTR
I	Switching - EXER Working: no switch Protection: no switch	EXER

5.3.3. State Transitions When Local Request Is Applied

In the state description below, 'O' means that a new local request will be rejected because of an existing request.

Initial state	New request	New state
A (Idle)	LP	C (Switching - LP)
	LW	D (Idle - LW)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	Recover from SF	N/A
	MS	G (Switching - MS)
	Clear	N/A
	WTR expires	N/A
	EXER	I (Switching - EXER)
Initial state	New request	New state
B (Pass-through)	LP	C (Switching - LP)
	LW	B (Pass-through)
	FS	O - if current state is due to LP sent by another node E (Switching - FS) - otherwise
	SF	O - if current state is due to LP sent by another node F (Switching - SF) - otherwise
	Recover from SF	N/A
	MS	O - if current state is due to LP, SF, or FS sent by another node G (Switching - MS) - otherwise
	Clear	N/A
	WTR expires	N/A
	EXER	O

Initial state	New request	New state
C (Switching - LP)	LP	N/A
	LW	O
	FS	O
	SF	O
	Recover from SF	N/A
	MS	O
	Clear	A (Idle) - if there is no failure in the ring
		F (Switching - SF) - if there is a failure at this node
		B (Pass-through) - if there is a failure at another node
	WTR expires	N/A
	EXER	O

Initial state	New request	New state
D (Idle - LW)	LP	C (Switching - LP)
	LW	N/A - if on the same link
		D (Idle - LW) - if on another link
	FS	O - if on the same link
		E (Switching - FS) - if on another link
	SF	O - if on the addressed link
		F (Switching - SF) - if on another link
	Recover from SF	N/A
	MS	O - if on the same link
		G (Switching - MS) - if on another link
	Clear	A (Idle) - if there is no failure on addressed link
		F (Switching - SF) - if there is a failure on this link
	WTR expires	N/A
	EXER	O

Initial state	New request	New state
E (Switching - FS)	LP	C (Switching - LP)
	LW	O - if on another link
		D (Idle - LW) - if on the same link
	FS	N/A - if on the same link
		E (Switching - FS) - if on another link
	SF	O - if on the addressed link
		E (Switching - FS) - if on another link
	Recover from SF	N/A
	MS	O
	Clear	A (Idle) - if there is no failure in the ring
		F (Switching - SF) - if there is a failure at this node
		B (Pass-through) - if there is a failure at another node
	WTR expires	N/A
	EXER	O

Initial state	New request	New state
F (Switching - SF)	LP	C (Switching - LP)
	LW	O - if on another link
		D (Idle - LW) - if on the same link
	FS	E (Switching - FS)
	SF	N/A - if on the same link
		F (Switching - SF) - if on another link
	Recover from SF	H (Switching - WTR)
	MS	O
	Clear	N/A
	WTR expires	N/A
	EXER	O

Initial state	New request	New state
G (Switching - MS)	LP	C (Switching - LP)
	LW	O - if on another link
		D (Idle - LW) - if on the same link
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	Recover from SF	N/A
	MS	N/A - if on the same link
		G (Switching - MS) - if on another link, release the switches but signal MS
	Clear	A
	WTR expires	N/A
	EXER	O

Initial state	New request	New state
H (Switching - WTR)	LP	C (Switching - LP)
	LW	D (Idle - W)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	Recover from SF	N/A
	MS	G (Switching - MS)
	Clear	A
	WTR expires	A
	EXER	O

Initial state	New request	New state
I (Switching - EXER)	LP	C (Switching - LP)
	LW	D (Idle - W)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	Recover from SF	N/A
	MS	G (Switching - MS)
	Clear	A
	WTR expires	N/A
	EXER	N/A - if on the same link
		I (Switching - EXER)

5.3.4. State Transitions When Remote Request is Applied

The priority of a remote request does not depend on the side from which the request is received.

Initial state	New request	New state
A (Idle)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	MS	G (Switching - MS)
	WTR	N/A
	EXER	I (Switching - EXER)
	RR	N/A
	NR	A (Idle)

Initial state	New request	New state
B (Pass-through)	LP	C (Switching - LP)
	FS	N/A - cannot happen when there is an LP request in the ring
		E (Switching - FS) - otherwise
	SF	N/A - cannot happen when there is an LP request in the ring
		F (Switching - SF) - otherwise
	MS	N/A - cannot happen when there is an LP, FS, or SF request in the ring
		G (Switching - MS) - otherwise
	WTR	N/A - cannot happen when there is an LP, FS, SF, or MS request in the ring
	EXER	N/A - cannot happen when there is an LP, FS, SF, MS, or a WTR request in the ring
		I (Switching - EXER) - otherwise
	RR	N/A
	NR	A (Idle) - if received from both sides

Initial state	New request	New state
C (Switching - LP)	LP	C (Switching - LP)
	FS	N/A - cannot happen when there is an LP request in the ring
	SF	N/A - cannot happen when there is an LP request in the ring
	MS	N/A - cannot happen when there is an LP request in the ring
	WTR	N/A
	EXER	N/A - cannot happen when there is an LP request in the ring
	RR	C (Switching - LP)
	NR	N/A

Initial state	New request	New state
D (Idle - LW)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	MS	G (Switching - MS)
	WTR	N/A
	EXER	I (Switching - EXER)
	RR	N/A
	NR	D (Idle - LW)

Initial state	New request	New state
E (Switching - FS)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	E (Switching - FS)
	MS	N/A - cannot happen when there is an FS request in the ring
	WTR	N/A
	EXER	N/A - cannot happen when there is an FS request in the ring
	RR	E (Switching - FS)
	NR	N/A

Initial state	New request	New state
F (Switching - SF)	LP	C (Switching - LP)
	FS	F (Switching - SF)
	SF	F (Switching - SF)
	MS	N/A - cannot happen when there is an SF request in the ring
	WTR	N/A
	EXER	N/A - cannot happen when there is an SF request in the ring
	RR	F (Switching - SF)
	NR	N/A

Initial state	New request	New state
G (Switching - MS)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	MS	G (Switching - MS) - release the switches but signal MS
	WTR	N/A
	EXER	N/A - cannot happen when there is an MS request in the ring
	RR	G (Switching - MS)
	NR	N/A

Initial state	New request	New state
H (Switching - WTR)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	MS	G (Switching - MS)
	WTR	H (Switching - WTR)
	EXER	N/A - cannot happen when there is a WTR request in the ring
	RR	H (Switching - WTR)
	NR	N/A

Initial state	New request	New state
I (Switching - EXER)	LP	C (Switching - LP)
	FS	E (Switching - FS)
	SF	F (Switching - SF)
	MS	G (Switching - MS)
	WTR	N/A
	EXER	I (Switching - EXER)
	RR	I (Switching - EXER)
	NR	N/A

5.3.5. State Transitions When Request Addresses to Another Node is Received

The priority of a remote request does not depend on the side from which the request is received.

Initial state	New request	New state
A (Idle)	LP	B (Pass-through)
	FS	B (Pass-through)
	SF	B (Pass-through)
	MS	B (Pass-through)
	WTR	B (Pass-through)
	EXER	B (Pass-through)
	RR	N/A
	NR	N/A

Initial state	New request	New state
B (Pass-through)	LP	B (Pass-through)
	FS	N/A - cannot happen when there is an LP request in the ring
		B (Pass-through) - otherwise
	SF	N/A - cannot happen when there is an LP request in the ring
		B (Pass-through) - otherwise
	MS	N/A - cannot happen when there is an LP, FS, or SF request in the ring
		B (Pass-through) - otherwise
	WTR	N/A - cannot happen when there is an LP, FS, SF, or MS request in the ring
		B (Pass-through) - otherwise
	EXER	N/A - cannot happen when there is an LP, FS, SF, MS, or a WTR request in the ring
		B (Pass-through) - otherwise
	RR	N/A
	NR	N/A

Initial state	New request	New state
C (Switching - LP)	LP	C (Switching - LP)
	FS	N/A - cannot happen when there is an LP request in the ring
	SF	N/A - cannot happen when there is an LP request in the ring
	MS	N/A - cannot happen when there is an LP request in the ring
	WTR	N/A - cannot happen when there is an LP request in the ring
	EXER	N/A - cannot happen when there is an LP request in the ring
	RR	N/A
	NR	N/A

Initial state	New request	New state
D (Idle - LW)	LP	B (Pass-through)
	FS	B (Pass-through)
	SF	B (Pass-through)
	MS	B (Pass-through)
	WTR	B (Pass-through)
	EXER	B (Pass-through)
	RR	N/A
	NR	N/A

Initial state	New request	New state
E (Switching - FS)	LP	B (Pass-through)
	FS	E (Switching - FS)
	SF	E (Switching - FS)
	MS	N/A - cannot happen when there is an FS request in the ring
	WTR	N/A - cannot happen when there is an FS request in the ring
	EXER	N/A - cannot happen when there is an FS request in the ring
	RR	N/A
	NR	N/A

Initial state	New request	New state
F (Switching - SF)	LP	B (Pass-through)
	FS	F (Switching - SF)
	SF	F (Switching - SF)
	MS	N/A - cannot happen when there is an SF request in the ring
	WTR	N/A - cannot happen when there is an SF request in the ring
	EXER	N/A - cannot happen when there is an SF request in the ring
	RR	N/A
	NR	N/A

Initial state	New request	New state
G (Switching - MS)	LP	B (Pass-through)
	FS	B (Pass-through)
	SF	B (Pass-through)
	MS	G (Switching - MS) - release the switches but signal MS
	WTR	N/A - cannot happen when there is an MS request in the ring
	EXER	N/A - cannot happen when there is an MS request in the ring
	RR	N/A
	NR	N/A

Initial state	New request	New state
H (Switching - WTR)	LP	B (Pass-through)
	FS	B (Pass-through)
	SF	B (Pass-through)
	MS	B (Pass-through)
	WTR	N/A
	EXER	N/A - cannot happen when there is a WTR request in the ring
	RR	N/A
	NR	N/A

Initial state	New request	New state
I (Switching - EXER)	LP	B (Pass-through)
	FS	B (Pass-through)
	SF	B (Pass-through)
	MS	B (Pass-through)
	WTR	N/A
	EXER	I (Switching - EXER)
	RR	N/A
	NR	N/A

6. IANA Considerations

IANA has assigned the values listed in the sections below.

6.1. G-ACh Channel Type

The Channel Types for G-ACh are allocated from the PW Associated Channel Type registry defined in [RFC4446] and updated by [RFC5586].

IANA has allocated the following new G-ACh Channel Type in the "MPLS Generalized Associated Channel (G-ACh) Types (including Pseudowire Associated Channel Types)" registry:

Value	Description	Reference
0x002A	Ring Protection Switching (RPS) Protocol	this document

6.2. RPS Request Codes

IANA has created the subregistry "MPLS RPS Request Code Registry" under the "Generic Associated Channel (G-ACh) Parameters" registry. All code points within this registry shall be allocated according to the "Specification Required" procedure as specified in [RFC8126].

The RPS request field is 8 bits; the allocated values are as follows:

Value	Description	Reference
0	No Request (NR)	this document
1	Reverse Request (RR)	this document
2	Unassigned	
3	Exercise (EXER)	this document
4	Unassigned	
5	Wait-to-Restore (WTR)	this document
6	Manual Switch (MS)	this document
7-10	Unassigned	
11	Signal Fail (SF)	this document
12	Unassigned	
13	Forced Switch (FS)	this document
14	Unassigned	
15	Lockout of Protection (LP)	this document
16-254	Unassigned	
255	Reserved	

7. Operational Considerations

This document describes three protection modes of the RPS protocol. Operators could choose the appropriate protection mode according to their network and service requirement.

Wrapping mode provides a ring protection mechanism in which the protected traffic will reach every node of the ring and is applicable to protect both the point-to-point LSPs and LSPs that need to be dropped in several ring nodes, i.e., the point-to-multipoint applications. When protection is inactive, the protected traffic is switched (wrapped) to/from the protection ring tunnel at both sides of the defective link/node. Due to the wrapping, the additional propagation delay and bandwidth consumption of the protection tunnel are considerable. For bidirectional LSPs, the protected traffic in both directions is co-routed.

Short-wrapping mode provides a ring protection mechanism that can be used to protect only point-to-point LSPs. When protection is inactive, the protected traffic is wrapped to the protection ring tunnel at the defective link/node and leaves the ring when the protection ring tunnel reaches the egress node. Compared with the wrapping mode, short-wrapping can reduce the propagation latency and bandwidth consumption of the protection tunnel. However, the two directions of a protected bidirectional LSP are not totally co-routed.

Steering mode provides a ring protection mechanism that can be used to protect only point-to-point LSPs. When protection is inactive, the protected traffic is switched to the protection ring tunnel at the ingress node and leaves the ring when the protection ring tunnel reaches the egress node. The steering mode has the least propagation delay and bandwidth consumption of the three modes, and the two directions of a protected bidirectional LSP can be kept co-routed.

Note that only one protection mode can be provisioned in the whole ring for all protected traffic.

8. Security Considerations

MPLS-TP is a subset of MPLS, thus it builds upon many of the aspects of the security model of MPLS. Please refer to [RFC5920] for generic MPLS security issues and methods for securing traffic privacy and integrity.

The RPS message defined in this document is used for protection coordination on the ring; if it is injected or modified by an attacker, the ring nodes might not agree on the protection action,

and the improper protection-switching action may cause a temporary break to services traversing the ring. It is important that the RPS message is used within a trusted MPLS-TP network domain as described in [RFC6941].

The RPS message is carried in the G-ACh [RFC5586], so it is dependent on the security of the G-ACh itself. The G-ACh is a generalization of the Associated Channel defined in [RFC4385]. Thus, this document relies on the security mechanisms provided for the Associated Channel as described in those two documents.

As described in the security considerations of [RFC6378], the G-ACh is essentially connection oriented, so injection or modification of control messages requires the subversion of a transit node. Such subversion is generally considered hard in connection-oriented MPLS networks and impossible to protect against at the protocol level. Management-level techniques are more appropriate. The procedures and protocol extensions defined in this document do not affect the security model of MPLS-TP linear protection as defined in [RFC6378].

9. References

9.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <https://www.rfc-editor.org/info/rfc3031>.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, DOI 10.17487/RFC4385, February 2006, <https://www.rfc-editor.org/info/rfc4385>.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, DOI 10.17487/RFC4446, April 2006, <https://www.rfc-editor.org/info/rfc4446>.

[RFC5586] Bocci, M., Ed., Vigoureux, M., Ed., and S. Bryant, Ed., "MPLS Generic Associated Channel", RFC 5586, DOI 10.17487/RFC5586, June 2009, <https://www.rfc-editor.org/info/rfc5586>.

[RFC5654] Niven-Jenkins, B., Ed., Brungard, D., Ed., Betts, M., Ed., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, DOI 10.17487/RFC5654, September 2009, <https://www.rfc-editor.org/info/rfc5654>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

9.2. Informative References

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, DOI 10.17487/RFC5920, July 2010, <https://www.rfc-editor.org/info/rfc5920>.

[RFC6371] Busi, I., Ed. and D. Allan, Ed., "Operations, Administration, and Maintenance Framework for MPLS-Based Transport Networks", RFC 6371, DOI 10.17487/RFC6371, September 2011, <https://www.rfc-editor.org/info/rfc6371>.

[RFC6378] Weingarten, Y., Ed., Bryant, S., Osborne, E., Sprecher, N., and A. Fulignoli, Ed., "MPLS Transport Profile (MPLS-TP) Linear Protection", RFC 6378, DOI 10.17487/RFC6378, October 2011, <https://www.rfc-editor.org/info/rfc6378>.

[RFC6941] Fang, L., Ed., Niven-Jenkins, B., Ed., Mansfield, S., Ed., and R. Graveman, Ed., "MPLS Transport Profile (MPLS-TP) Security Framework", RFC 6941, DOI 10.17487/RFC6941, April 2013, <https://www.rfc-editor.org/info/rfc6941>.

[RFC6974] Weingarten, Y., Bryant, S., Ceccarelli, D., Caviglia, D., Fondelli, F., Corsi, M., Wu, B., and X. Dai, "Applicability of MPLS Transport Profile for Ring Topologies", RFC 6974, DOI 10.17487/RFC6974, July 2013, <https://www.rfc-editor.org/info/rfc6974>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

Acknowledgements

The authors would like to thank Gregory Mirsky, Yimin Shen, Eric Osborne, Spencer Jackson, and Eric Gray for their valuable comments and suggestions.

Contributors

The following people contributed significantly to the content of this document and should be considered co-authors:

Kai Liu
Huawei Technologies
Email: alex.liukai@huawei.com

Jia He
Huawei Technologies
Email: hejia@huawei.com

Fang Li
China Academy of Telecommunication Research MIIT
China
Email: lifang@catr.cn

Jian Yang
ZTE Corporation
China
Email: yang.jian90@zte.com.cn

Junfang Wang
Fiberhome Telecommunication Technologies Co., LTD.
Email: wjf@fiberhome.com.cn

Wen Ye
China Mobile
Email: yewen@chinamobile.com

Minxue Wang
China Mobile
Email: wangminxue@chinamobile.com

Sheng Liu
China Mobile
Email: liusheng@chinamobile.com

Guanghui Sun
Huawei Technologies
Email: sunguanghui@huawei.com

Authors' Addresses

Weiqiang Cheng
China Mobile

Email: chengweiqiang@chinamobile.com

Lei Wang
China Mobile

Email: wangleiyj@chinamobile.com

Han Li
China Mobile

Email: lihan@chinamobile.com

Huub van Helvoort
Hai Gaoming BV

Email: huubatwork@gmail.com

Jie Dong
Huawei Technologies

Email: jie.dong@huawei.com